

## Respondent-Driven Sampling of Injection Drug Users in Two U.S.–Mexico Border Cities: Recruitment Dynamics and Impact on Estimates of HIV and Syphilis Prevalence

Simon D. W. Frost, Kimberly C. Brouwer, Michelle A. Firestone Cruz, Rebeca Ramos, Maria Elena Ramos, Remedios M. Lozada, Carlos Magis-Rodriguez, and Steffanie A. Strathdee

**ABSTRACT** Respondent-driven sampling (RDS), a chain referral sampling approach, is increasingly used to recruit participants from hard-to-reach populations, such as injection drug users (IDUs). Using RDS, we recruited IDUs in Tijuana and Ciudad (Cd.) Juárez, two Mexican cities bordering San Diego, CA and El Paso, TX, respectively, and compared recruitment dynamics, reported network size, and estimates of HIV and syphilis prevalence. Between February and April 2005, we used RDS to recruit IDUs in Tijuana (15 seeds, 207 recruits) and Cd. Juárez (9 seeds, 197 recruits), Mexico for a cross-sectional study of behavioral and contextual factors associated with HIV, HCV and syphilis infections. All subjects provided informed consent, an anonymous interview, and a venous blood sample for serologic testing of HIV, HCV, HBV (Cd. Juárez only) and syphilis antibody. Log-linear models were used to analyze the association between the state of the recruiter and that of the recruitee in the referral chains, and population estimates of the presence of syphilis antibody were obtained, correcting for biased sampling using RDS-based estimators. Sampling of the targeted 200 recruits per city was achieved rapidly (2 months in Tijuana, 2 weeks in Cd. Juárez). After excluding seeds and missing data, the sample prevalence of HCV, HIV and syphilis were 96.6, 1.9 and 13.5% respectively in Tijuana, and 95.3, 4.1, and 2.7% respectively in Cd. Juárez (where HBV prevalence was 84.7%). Syphilis cases were clustered in recruitment trees. RDS-corrected estimates of syphilis antibody prevalence ranged from 12.8 to 26.8% in Tijuana and from 2.9 to 15.6% in Ciudad Juárez, depending on how recruitment patterns were modeled, and assumptions about how network size affected an individual's probability of being included in the sample. RDS was an effective method to rapidly recruit IDUs in these cities. Although the frequency of HIV was low, syphilis prevalence was high, particularly in Tijuana. RDS-corrected estimates of syphilis prevalence were sensitive to model assumptions, suggesting that further validation of RDS is necessary.

**KEYWORDS** HIV and syphilis prevalence, Injection drug users, Respondent driven sampling.

Frost is with the Department of Pathology and Antiviral Research Center, University of California, San Diego, CA, USA; Brouwer, Firestone Cruz, and Strathdee are with the Division of International Health and Cross Cultural Medicine, Department of Family and Preventive Medicine, University of California, San Diego, CA, USA; R. Ramos is with the United States–Mexico Border Health Association, El Paso, TX, USA; M. E. Ramos is with Programa Compañeros, Ciudad Juárez, Chihuahua, Mexico; Lozada is with Patronato COMUSIDA, Tijuana, Baja California, Mexico; Magis-Rodriguez is with the Centro Nacional para la Prevención y el Control del VIH/SIDA (CENSIDA), México City, México.

Correspondence: Simon D. W. Frost, UCSD Antiviral Research Center, 150 W. Washington St., Suite 100, San Diego, CA 92103, USA. (E-mail: [sdfrost@ucsd.edu](mailto:sdfrost@ucsd.edu))

## INTRODUCTION

Injection drug use is a growing problem in cities located along the U.S.–Mexico border. Approximately 70% of U.S. cocaine originating in South America passes through the Central America–Mexico corridor.<sup>1</sup> Cities and towns positioned on drug trafficking routes often experience epidemics of injection drug use.<sup>2,3</sup> Injection drug users (IDUs) are at high risk of blood-borne infections, such as hepatitis C virus (HCV) and human immunodeficiency virus type-1 (HIV-1) infection, and of acquiring HIV-1 and other sexually transmitted infections (STIs) through high rates of unprotected sex.<sup>4–10</sup>

Mexico is currently considered a country of low HIV/AIDS prevalence<sup>11</sup> (180,000 adult cases in 2005, a seroprevalence in the general population of 0.3%<sup>12</sup>), and the HIV epidemic has been mainly confined to men who have sex with men.<sup>13</sup> Although injection drug use appears to have played only a minor role in the epidemic on a country-wide level,<sup>14</sup> injection drug use appears to be increasingly important as a risk factor for HIV infection in some Mexican cities bordering the U.S. Viani et al.<sup>15</sup> noted that the prevalence of HIV among pregnant women giving birth at Tijuana General Hospital rose from 0.29% in 1998 to 1.02% in 2001 and, in a subsequent study, showed that pregnant HIV-infected women were more likely to either inject drugs or to have a spouse/partner who injected drugs.<sup>16</sup> In 2002, Valdez et al.<sup>17</sup> reported that 21% of female sex workers in Ciudad (Cd.) Juárez injected illicit drugs, whereas a study by Patterson et al.<sup>18</sup> in 2005 showed that over half of female sex workers in Cd. Juárez injected drugs, suggesting increasing overlap between sexual and IDU networks.

Overlap between injection drug use and the trade of sex for money or drugs may contribute to elevated risk of STIs other than HIV, such as syphilis. Syphilis has been associated with higher HIV seroprevalence in a number of populations and is considered a cofactor of HIV transmission.<sup>19–22</sup> In contrast to HIV, syphilis has been present in Mexico since at least the time of the Spanish invasion; however, the number of reported cases has decreased from 40,607 in 1945 (190.5 per 100,000) to 2,608 in 1990 (3.2 per 100,000).<sup>23</sup>

Aggregate figures for syphilis prevalence belie the sub-epidemics occurring within specific risk groups. Several studies have been conducted with female sex workers in Mexico and have found varying syphilis prevalence levels. In 1990, 23.7% of 1,386 sex workers in four Mexican states had a reactive syphilis test.<sup>24</sup> In 1993, testing of 826 sex workers in Mexico City showed an overall prevalence of 6.4%, with different syphilis rates associated with different patterns of sex work: 1.3% for massage parlor workers, 4.4% for bar girls, and 9.6% for streetwalkers.<sup>25</sup> The prevalence of syphilis among 3,100 female sex workers tested at an AIDS clinic during 1992 and 1993 was 8.2%.<sup>26</sup> In contrast, syphilis prevalence was low (2.3 and 1.1%) among gynecological outpatients in two Mexican cities between the years 1994 and 1995.<sup>27</sup> However, little is known about syphilis prevalence in IDU populations in Mexico.

In order to estimate the prevalence of HIV, HCV, and syphilis among IDUs, we conducted a cross-sectional study of IDUs in the border cities of Tijuana and Cd. Juárez, Mexico. Both cities are located on major drug trafficking routes and have large IDU populations (c. 6,000), with a similar sex ratio among the IDUs (c. 80% male).<sup>3</sup> As stigma surrounding injection drug use makes it difficult to obtain a representative sample of injection drug users, we recruited individuals using respondent-driven sampling (RDS).<sup>28,29</sup> By collecting data on individuals' personal network sizes, RDS attempts to correct for biases in the sampling process, in order

to obtain unbiased estimates of parameters such as the prevalence of a disease. In this study, we report on patterns of recruitment and the prevalence of HIV, HCV, and syphilis (and hepatitis B infection, for Cd. Juárez) in the context of sexual risk.

## MATERIALS AND METHODS

### Study Population

From February through April 2005, IDUs were enrolled in a cross-sectional study in Tijuana and Cd. Juárez, Mexico. Eligibility criteria for the study included: having injected illicit drugs within the past month, confirmed by inspection of injection stigmata ('track marks'); aged 18 years or older; willing and able to provide informed consent; and not having been previously interviewed for the study. Subjects gave their written informed consent to participate in the study. Study methods were approved by the Institutional Review Board of the University of California, San Diego and the Ethics Board of the Tijuana General Hospital, which has one of the few federal-wide assurances in Mexico.

### Recruitment

RDS methods were used to recruit participants.<sup>28,29</sup> A diverse group of "seeds" (heterogeneous in age, gender, and geographic location) were selected to initiate the process. After providing informed consent, seeds underwent an interview, were educated on how to refer other eligible IDUs, and were given three uniquely coded coupons to refer their peers. Coupons were given to participants until approximately 150 participants were recruited in order to obtain a target sample size of approximately 200 per site.

On each coupon, the study name, locations where they could participate, and a brief explanation was printed. In Cd. Juárez, interviews were conducted at a clinic run by Programa Compañeros, A.C., which is a trusted and well-respected non-governmental organization (NGO) that has been providing services to and conducting studies with IDUs in the city for decades. In Tijuana, staff from both COMUSIDA, the municipal HIV/AIDS program, and the Centro de Integración y Recuperación para Enfermos de Alcoholismo y Drogadicción "Mario Camacho Espíritu", A.C. (CIRAD), an NGO that began working with drug users in 1991, made weekly trips to three geographically diverse 'colonias' (i.e., neighborhoods) in the city: Zona Norte, Grupo México, and Sepanal, using a modified recreational vehicle that operated as a mobile clinic (the 'Prevemovihl').

Monetary reimbursements were given to participants to cover transportation costs and to compensate them for their time. The study staff in each site proposed the incentive levels based on their experience with this population and the incentives for previous studies. Participants in Cd. Juárez received \$20 U.S. dollars (USD) for participation in the baseline visit and \$5 USD when receiving laboratory test results at a one month follow-up visit. In Tijuana, \$10 USD was given at baseline and \$5 for the follow-up visit. In addition, participants at both sites were given \$5 for each eligible person they recruited. These levels were not regarded as high.

### Data Collection

Upon enrollment, trained staff administered quantitative surveys eliciting information on topics such as socio-economic and demographic profiles, drug use practices,

sterile syringe access, barriers to sterile syringe use, experience with drug abuse treatment and incarceration history, health status, and HIV knowledge and testing history. We also asked about sexual behaviors and condom use with regular, casual and client partners of the opposite and same sex. Questions pertained to lifetime risk behaviors and those occurring in the prior six months.

For RDS purposes, we measured network size using the question “En los últimos 6 meses, ¿cuántas personas conoce de nombre o de apodo que se han inyectado drogas?” (“In the past 6 months, how many people do you know by name or street name who have injected drugs?”). To determine the relationship between recruiter and recruitee, we asked “¿Cuál es su relación con la persona que le entregó el cupón?” (“What is your relationship to the person who gave you the coupon?”). Participants were given the choice of: “parientes” (relative); “pareja sexual” (sex partner/spouse); “amigo(a)” (friend); “conocido” (acquaintance); “desconocido” (stranger); and “otro” (other). To determine the size of individuals’ networks with respect to injection drug use, we asked “En los últimos 6 meses, ¿con cuántas personas diferentes acostumbra inyectarse?” (“In the last 6 months, on average how many different people did you usually inject with?”). After the interview, blood was drawn for antibody testing of HIV, HCV, HBV (Cd. Juárez only), and syphilis. Pre- and post-test counseling, and referral to treatment where indicated, was provided to all participants.

### Laboratory Samples

Blood samples were obtained by venipuncture and serum was stored at the municipal health clinic in Tijuana or Cd. Juárez before being shipped frozen to the San Diego County public health laboratory or New Mexico State Laboratory, respectively. All participants were screened on-site in Mexico for HIV with the Determine rapid test (Abbott Laboratories). For the Tijuana samples, in the event of an HIV-positive or indeterminate test, results were confirmed with a Western blot, HIV enzyme immunoassay (EIA), and HIV immunofluorescence assay. For samples from Cd. Juárez, the HIV EIA was conducted on all samples, and a confirmatory Western blot was performed on positive or indeterminate sample. Cd. Juárez samples were also tested for hepatitis B antigen (Genetic Systems HBsAg EIA 3.0, Bio-Rad Laboratories) and antibody (DiaSorin ETI-AB-COREK PLUS). All samples were tested for syphilis with the rapid plasma reagin (RPR) test (Macro-Vue, Becton Dickinson) and if reactive, confirmed by a *Treponema pallidum* particle agglutination assay (TPPA; Fujirebio Diagnostics).

### Statistical Methods

Obtaining estimates of population proportions of groups using RDS involves combining three kinds of data: the sample proportion of each group, the crosstabulation of groups between pairs of recruiters and recruitees, and differences in network size between groups. To estimate equilibrium proportions of different groups, and to estimate the pattern of mixing between groups, we assumed that the recruitment process followed a first order Markov process.<sup>28,29</sup> Under this model, the relationship between the state of the recruiter and recruitee can be modeled using log-linear models applied to a two-way table of counts.<sup>30,31</sup> We classified individuals by sex and syphilis seropositivity and fitted a series of hierarchical log-linear models of increasing complexity to the data to determine patterns of nonrandom mixing between groups along each recruitment tree, choosing the best model as that which had the lowest value of Akaike’s Information Criterion.<sup>32</sup> For

the purposes of analysis, we considered all individuals with positive syphilis tests based upon RPR, and did not classify individuals further into those with TPPA titers greater than or equal to 1:8 (who may represent infectious cases) and those with titers of 1:1 to 1:4 (who may represent past infection).

To derive RDS-corrected estimates of syphilis seropositivity in men and women in Tijuana and Ciudad Juárez, we estimated recruitment weights for each group (as the ratio of the equilibrium to sample proportions of each group). We estimated the equilibrium fraction as previously described.<sup>29</sup> We used both raw counts and predicted counts based on the best fitting log-linear model. Degree weights were estimated using linear least squares.<sup>29</sup> We used both unadjusted and adjusted estimates of personal network size.<sup>33</sup> An overall sampling weight was derived for each group, from which population-level estimates were obtained.

Pre-processing of the data was performed using Stata v. 8.2 (Stata Corporation, College Station, TX). Networks and trees were generated using scripts written in Python and visualized using GraphViz (AT&T Research, Florham Park, NJ). Statistical analyses and summary statistics of the recruitment network were generated in R,<sup>34</sup> and RDS based corrections were calculated using Maxima (<http://maxima.sourceforge.net>). We chose to develop our own programs rather than use RDSAT (<http://www.respondentdrivensampling.org>) primarily to familiarize ourselves with the statistical theory underlying RDS-based corrections. All code is available from the first author on request.

## RESULTS

### Study Population

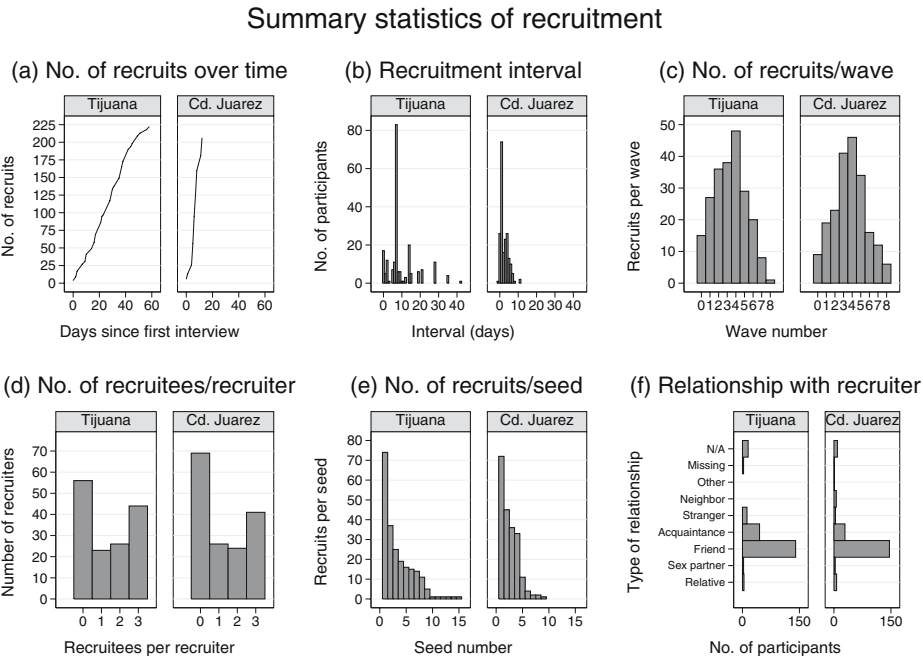
Table 1 summarizes some basic data relating to the Tijuana (15 seeds, 207 recruits) and Cd. Juárez (9 seeds, 197 recruits) study populations. Both populations were predominantly male, with participants in their early to mid-30s. Crude HCV seroprevalence was extremely high (>95%) in both cities. Hepatitis B seroprevalence was only determined for Cd. Juárez, where it was high (85% overall); only one individual was positive for HBV antigen. Crude HIV seroprevalence was low, but the crude prevalence of syphilis was high, especially among women.

### Recruitment Dynamics

RDS was an effective means of recruiting IDUs in both cities. The number of individuals recruited increased rapidly following the first interview, especially in Cd. Juárez (Figure 1a), where many individuals interviewed the same day as their recruiter (Figure 1b). Apart from these differences in the tempo of recruitment, patterns of recruitment were very similar between the two cities; recruitment was highest in the fourth wave of recruitment, with some individuals being recruited after eight waves, suggesting that despite rapid recruitment, good sociometric depth was obtained (Figure 1c). After excluding individuals who were not given coupons, the number of recruits per recruiter showed a bimodal distribution, with many individuals either recruiting zero or three recruits (Figure 1d), suggesting the presence of a mixed population of ineffective and effective recruiters. In both cities, approximately one half of participants were recruited via referral trees originating from two seeds (Figure 1e). The relationship between recruiter and recruit was usually ‘friend,’ ‘acquaintance,’ or another close relationship such as a family

**TABLE 1. Summary statistics of age, parameters pertaining to risk of STI, and seroprevalence of HIV, HCV and syphilis, by city, sex, and by whether individuals were seeds, or recruits**

City	Type of recruit	Tijuana				Ciudad Juárez			
		Seeds		Recruits		Seeds		Recruits	
Gender (n)		Male (9)	Female (6)	Male (194)	Female (13)	Male (6)	Female (3)	Male (184)	Female (13)
Age (median, range)		35 (29–54)	28 (22–32)	34 (18–56)	32 (24–48)	42 (33–49)	33 (30–42)	33 (18–62)	36 (20–48)
Age at sexual debut (median, range)		15 (12–18)	16.5 (9–27)	15 (7–28) [n = 191]	14 (11–15)	13.5 (8–18)	16 (15–18)	15 (5–23) [n = 181]	14 (12–18)
No. of lifetime male sexual partners (median, range)		2 (0–15)	8 (0–10)	0 (0–100) [n = 186]	20 (2–500) [n = 11]	0 (0–2)	5 (3–6)	0 (0–15) [n = 178]	5 (1–100) [n = 11]
No. of lifetime female sexual partners (median, range)		10 (4–300)	0 (0–1)	10 (0–500) [n = 191]	0 (0–4)	22 (8–200)	5 (0–10) [n = 2]	10 (1–300) [n = 175]	0 (0–7) [n = 12]
Prostitution as main source of income over the last 6 months (fraction, %)		0/9 (0)	1/6 (16.7)	1/191 (0.5)	6/13 (46.2)	0/6 (0)	1/3 (33)	0/184 (0)	1/13 (7.7)
Had sex in last 6 months (fraction, %)		5/9 (55.6)	4/6 (66.7)	106/194 (54.6)	12/13 (92.3)	4/6 (66.7)	2/3 (66.7)	108/184 (58.7)	10/13 (76.9)
Ever been given (bought) sex in the last 6 months (fraction, %)		1/3 (33.3)	2/3 (66.7)	24/67 (35.8)	8/8 (100)	0/2 (0)	0/0 (N/A)	10/57 (17.5)	4/5 (80.0)
Ever given (sold) sex in the last 6 months (fraction, %)		1/3 (33.3)	2/3 (66.7)	17/67 (25.4)	2/8 (25)	1/2 (50)	0/0 (N/A)	23/57 (40.4)	3/5 (60.0)
Injected drugs with sex partner in last 6 months		1/9 (11.1)	2/6 (33.3)	10/194 (5.15)	6/13 (46.2)	0/6 (0)	1/3 (33.3)	15/184 (8.2)	3/13 (2.3)
Syphilis RPR + (fraction, %)		1/7 (14.3)	2/6 (33.3)	20/194 (10.3)	8/13 (61.5)	0/4 (0)	2/3 (66.7)	3/175 (1.71)	2/11 (18.2)
HIV antibody (fraction, %)		1/7 (14.3)	1/6 (16.7)	4/194 (2.1)	0/13 (0)	0/6 (0)	0/3 (0)	8/182 (4.4)	0/12 (0)
HCV antibody (fraction, %)		6/7 (85.7)	6/6 (100)	188/194 (96.9)	12/13 (92.3)	6/6 (100)	3/3 (100)	174/181 (96.1)	10/12 (83.3)
HBV antibody (fraction, %)		N/A	N/A	N/A	N/A	6/6 (100)	3/3 (100)	154/181 (85.1)	8/12 (66.7)



**FIGURE 1.** Summary of the dynamics of recruitment, by city. (a) The cumulative number of recruits over time. (b) The interval between the interview of the recruiter and that of their recruitee (omitting those individuals who did not recruit). (c) The number of recruits in each recruitment wave from the seed. (d) The number of recruits per recruiter (excluding individuals who were not given any coupons). (e) The number of recruits from each seed. (f) The relationship between recruiter and recruitee.

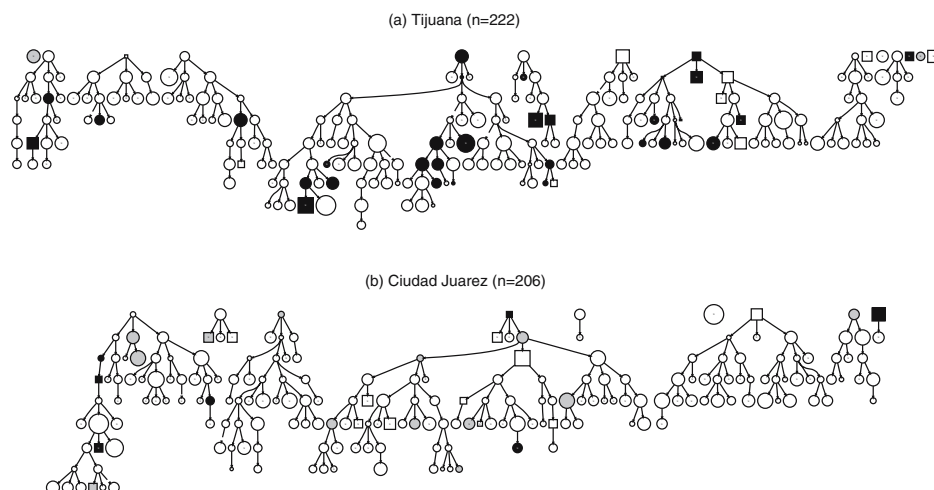
member or a sex partner (Figure 1f), which is important as RDS-based estimates assume that these relationships are reciprocal.<sup>28,29</sup>

**Recruitment Trees**

Figure 2 shows a ‘forest’ of recruitment trees for each city, with the syphilis antibody status indicated by shading and the gender of each individual indicated by different symbols. This figure illustrates that the sample prevalence of syphilis was higher in Tijuana than in Cd. Juárez and that syphilis cases appeared to cluster in the recruitment trees. Although there was a low frequency of women in the sample, a disproportionate number of women also had syphilis.

To test whether syphilis prevalence differed by sex and whether cases of syphilis were clustered, we analyzed the relationship between the sex and syphilis status of the recruiter and recruitee in Tijuana and Cd. Juárez (Table 2). A set of hierarchical log-linear models were used to test whether there was an association between syphilis seropositivity and sex and between the state of the recruiter and that of the recruitee, with the best fitting model chosen using Akaike’s Information Criterion, AIC,<sup>32</sup> with lower values indicating better fit (Table 3). This approach is functionally equivalent to calculating the RDS homophily index, for which 1 denotes a perfect association, 0 a zero association, and –1 a perfect negative association. The reason for using log-linear models is that they offer a solid statistical framework with which we can





**FIGURE 2.** Recruitment networks (strictly speaking, a forest of recruitment trees) for the RDS based samples of IDUs in (a) Tijuana and (b) Ciudad Juárez. Seeds are shown at the *top* of the figure, and *arrows* indicate the direction of recruitment. Syphilis serostatus is shown by *shading*: *black*-syphilis antibody positive, *white*- syphilis antibody negative, *gray*- missing data. The gender of participants is indicated by the shape of the symbol: *square* for female and *circle* for male. The size of the symbol is related to the reported network size: the larger the symbol, the larger the network size. *Symbols marked with an 'x'* denote individuals who were given coupons, but did not recruit.

compare different models of association between the characteristics of the recruiter and those of the recruitee.

For both samples, the best fitting model included a significant (positive) association between the syphilis serostatus of the recruiter and the syphilis serostatus of the recruitee and was not significantly different from a 'saturated' model, in which each cell in the table is modeled with a single parameter. For the Tijuana sample, there was a positive association between the sex of the recruiter and the sex of the recruit, and, independently of this association, syphilis antibody positive women were disproportionately less likely to recruit syphilis positive men than syphilis negative men.

### Network Size

There was great variation in both measures of network size, with the average number of known IDUs an order of magnitude higher than the number of injecting partners. There was no correlation in network size between recruiter and recruitee, measured using either the number of IDUs known (Spearman's  $\rho = 0.0328$  and  $0.0204$  for Tijuana and Cd. Juárez, respectively) or the number of injecting partners as a measure of network size (Spearman's  $\rho = -0.006$  and  $-0.0121$  for Tijuana and Cd. Juárez, respectively; Figure 3a, b). Network sizes were similar when the sample was grouped by syphilis serostatus and sex (Figure 3c, d). As chain referral samples are biased towards individuals with larger network sizes, we adjusted the distribution of network size by weighting the distribution of network sizes by the inverse of the network size.<sup>33</sup> This reweighting led to a significant drop in estimated network size, from a median of 25–50 to 4–5 in Tijuana and from 15–20 to 8–10 in Cd. Juárez (Figure 3e).



TABLE 2 Relationship between recruiter and recruitee in terms of sex and syphilis antibody status

		Tijuana				Ciudad Juárez			
From	To	Male		Female		Male		Female	
		RPR <sup>−</sup>	RPR <sup>+</sup>	RPR <sup>−</sup>	RPR <sup>+</sup>	RPR <sup>−</sup>	RPR <sup>+</sup>	RPR <sup>−</sup>	RPR <sup>+</sup>
Male	RPR <sup>−</sup>	145 (145.2)	15 (14.66)	2 (2.33)	5 (4.82)	147 (146.7)	3 (2.30)	7 (7.43)	1 (1.53)
	RPR <sup>+</sup>	15 (15.23)	5 (4.91)	1 (0.24)	1 (1.61)	1 (1.59)	0 (0.2)	0 (0.08)	1 (0.13)
Female	RPR <sup>−</sup>	8 (7.19)	0 (0.11)	1 (1.29)	0 (0.41)	7 (6.50)	0 (0.10)	0 (0.33)	0 (0.07)
	RPR <sup>+</sup>	6 (6.38)	0 (0.32)	1 (1.14)	2 (1.16)	3 (3.17)	0 (0.40)	1 (0.16)	0 (0.27)

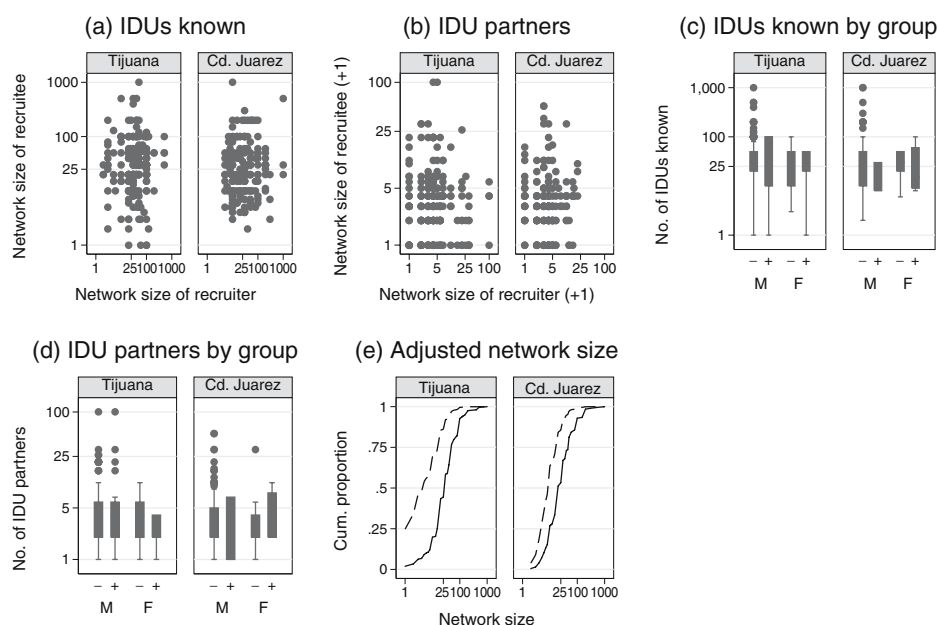
Smoothed estimates are shown in parentheses and are based upon the best fitting log linear model for each sample (Table 3).

TABLE 3. Fit of 12 log-linear models to the data shown in Table 2

Model	p	Tijuana				Ciudad Juárez			
		InL	AIC	LRT, G <sup>2</sup>	InL	AIC	LRT, G <sup>2</sup>	InL	LRT, G <sup>2</sup>
Symmetric (sex <sub>1</sub> , sex <sub>2</sub> ) + Symmetric (rpr <sub>1</sub> , rpr <sub>2</sub> )	5	−42.06	94.12	40.4 ( $p < 0.001$ )	−28.84	67.68	29.3 ( $p = 0.001$ )	−28.84	29.3 ( $p = 0.001$ )
sex <sub>1</sub> + rpr <sub>1</sub> + sex <sub>2</sub> + rpr <sub>2</sub>	5	−46.27	102.5	48.8 ( $p < 0.001$ )	−29.91	69.82	31.4 ( $p = 0.001$ )	−29.91	31.4 ( $p = 0.001$ )
sex <sub>1</sub> + rpr <sub>1</sub> + sex <sub>2</sub> * rpr <sub>2</sub>	6	−37.27	86.55	30.8 ( $p = 0.001$ )	−27.25	66.49	26.1 ( $p = 0.004$ )	−27.25	26.1 ( $p = 0.004$ )
sex <sub>1</sub> * rpr <sub>1</sub> + sex <sub>2</sub> * rpr <sub>2</sub>	7	−30.32	74.64	16.9 ( $p = 0.05$ )	−19.21	52.43	10.0 ( $p = 0.35$ )	−19.21	10.0 ( $p = 0.35$ )
sex <sub>1</sub> * rpr <sub>1</sub> + sex <sub>2</sub> * rpr <sub>2</sub> + sex <sub>1</sub> * sex <sub>2</sub>	8	−27.47	70.95	11.2 ( $p = 0.189$ )	−19.12	54.24	9.81 ( $p = 0.28$ )	−19.12	9.81 ( $p = 0.28$ )
sex <sub>1</sub> * rpr <sub>1</sub> + sex <sub>2</sub> * rpr <sub>2</sub> + rpr <sub>1</sub> * rpr <sub>2</sub>	8	−28.31	72.61	12.9 ( $p = 0.115$ )	−18.16	52.32	7.89 ( $p = 0.44$ )	−18.16	7.89 ( $p = 0.44$ )
sex <sub>1</sub> * rpr <sub>1</sub> + sex <sub>2</sub> * rpr <sub>2</sub> + sex <sub>1</sub> * sex <sub>2</sub> + rpr <sub>1</sub> * rpr <sub>2</sub>	9	−26.01	70.02	8.31 ( $p = 0.306$ )	−18.14	54.28	7.85 ( $p = 0.35$ )	−18.14	7.85 ( $p = 0.35$ )
sex <sub>1</sub> * rpr <sub>1</sub> + sex <sub>2</sub> * rpr <sub>2</sub> + sex <sub>1</sub> * sex <sub>2</sub> + rpr <sub>1</sub> * rpr <sub>2</sub> + sex <sub>1</sub> * rpr <sub>2</sub>	10	−23.87	67.73	4.02 ( $p = 0.674$ )	−16.48	52.96	4.53 ( $p = 0.61$ )	−16.48	4.53 ( $p = 0.61$ )
sex <sub>1</sub> * rpr <sub>1</sub> + sex <sub>2</sub> * rpr <sub>2</sub> + sex <sub>1</sub> * sex <sub>2</sub> + rpr <sub>1</sub> * rpr <sub>2</sub> + rpr <sub>1</sub> * sex <sub>2</sub>	10	−25.63	71.27	7.56 ( $p = 0.272$ )	−16.17	52.35	3.92 ( $p = 0.69$ )	−16.17	3.92 ( $p = 0.69$ )
sex <sub>1</sub> * rpr <sub>1</sub> + sex <sub>2</sub> * rpr <sub>2</sub> + sex <sub>1</sub> * sex <sub>2</sub> + rpr <sub>1</sub> * rpr <sub>2</sub> + sex <sub>1</sub> * rpr <sub>2</sub> + rpr <sub>1</sub> * sex <sub>2</sub>	11	−23.67	69.34	3.62 ( $p = 0.604$ )	−16.65	53.29	4.87 ( $p = 0.56$ )	−16.65	4.87 ( $p = 0.56$ )
Symmetric (sex <sub>1</sub> * rpr <sub>1</sub> , sex <sub>2</sub> * rpr <sub>2</sub> )	10	−25.91	71.82	8.1 ( $p = 0.23$ )	−15.03	52.06	1.63 ( $p = 0.90$ )	−15.03	1.63 ( $p = 0.90$ )
Saturated model (sex <sub>1</sub> * rpr <sub>1</sub> * sex <sub>2</sub> * rpr <sub>2</sub> )	16	−21.86	75.71	NA	−14.21	60.43	NA	−14.21	NA

The variables in the ‘Model’ column correspond to the sex of the recruiter (sex<sub>1</sub>) and recruitee (sex<sub>2</sub>) and the syphilis serostatus of the recruiter (rpr<sub>1</sub>) and recruitee (rpr<sub>2</sub>). The term ‘Symmetric(a, b)’ applies to a table which is symmetric about the diagonal of the table a by b. The number of parameters is denoted *p*, the log-likelihood of the model (higher is better) by InL, Akaike’s information Criterion (lower is better) by AIC, and a likelihood ratio test for goodness of fit against a saturated model by LRT. The best fitting model for each city, based on the lowest AIC, is indicated by underlining.

### Summary of network size



**FIGURE 3.** Summary of the network size distribution. (a) Scatterplot of the number of IDUs known by name or street name between recruiter and recruitee. (b) Scatterplot of the number of injecting partners between recruiter and recruitee. (c) Boxplot of the number of IDUs known by sex (*M* male, *F* female) and by syphilis antibody status (+ positive, – negative). (d) Boxplot of the number of injecting partners by sex (*M* male, *F* female) and by syphilis antibody status (+ positive, – negative). (e) Cumulative distribution of the number of IDUs known before (*solid line*) and after (*dashed line*) adjusting for biased sampling of individuals.

### Correcting for Sampling Bias

Information on network size, and on who recruited whom, collected as part of RDS, allows population estimates to be generated from the sample, despite biases in the sampling process. To do so, we need to estimate the pattern of mixing between different groups along the recruitment network and to determine how different the network size is in each group. The seroprevalence of HIV was too low, and the seroprevalence of HCV was too high in order to obtain meaningful correction factors; hence we concentrated on obtaining population estimates of the prevalence of syphilis antibody in men and women in the two cities.

In order to obtain unbiased population estimates of the prevalence of syphilis antibody, we used a poststratification process, in which we calculated recruitment weights and degree weights, which can be combined to give an overall sampling weight.<sup>46</sup> As RDS is a chain referral method, the prevalence of syphilis in the sample may have been different if recruitment had continued for further waves. Assuming that recruitment follows a first order Markov process, the ‘equilibrium prevalence’ of syphilis can be calculated from the cross-tabulations of the syphilis status of the recruiter and the recruitee. Using these estimates, we calculated recruitment weights, as the ratio of the equilibrium to sample frequencies of each group (Table 4); the closer these weights are to 1, the more representative the sampling of the group. For the Tijuana sample, recruitment weights were close to 1

for three groups, except for syphilis negative women, whose recruitment weight was 0.87 (i.e., this group was oversampled). For the Cd. Juárez sample, recruitment weights were close to 1, with the exception of syphilis positive women (1.25), who were undersampled.

Samples obtained using RDS can also be biased due to differences in network size between the groups. To compensate for this effect, we calculated degree weights, based upon the reported network size of known IDUs using the adjusted mean estimates of network size<sup>33</sup> in each group (Figure 3f). After calculating degree weights for each group, and multiplying them by recruitment weights to generate an overall sampling weight, the RDS corrected estimates of syphilis antibody prevalence were higher than those in the overall sample for both cities (Table 4).

### Sensitivity of RDS Estimates

To determine the sensitivity of point estimates of the prevalence of syphilis antibody to modeling assumptions, we also obtained estimates using unadjusted rather than adjusted network sizes and using a ‘smoothed’ transition matrix based on the best fitting log-linear model, rather than the raw counts. We found that RDS-based estimates were highly sensitive to these assumptions (Table 5). Estimated syphilis seroprevalence ranged from 12.4 to 26.8% in Tijuana and from 2.9 to 15.6% in Cd. Juárez, depending on how the pattern of recruitment was modeled and how reported network size was assumed to affect an individual’s probability of being included in the sample. However, our results suggest that syphilis seroprevalence is higher among women than men and higher in Tijuana than in Cd. Juárez and that sample proportions of syphilis using RDS in these populations may be underestimates of the true population seroprevalence.

### CONCLUSIONS

Respondent driven sampling offers the promise of a probability sample of individuals from hidden and hard-to-reach populations. RDS was originally developed in the context of recruiting IDUs<sup>35–39</sup> and, in our context, was an efficient method to recruit IDUs in two Mexican cities bordering the U.S. Recruitment was extremely rapid in Cd. Juárez compared to Tijuana, which may be due to greater access to the study site, higher monetary incentives, and that Programa Compañeros is more established in Cd. Juárez than CIRAD in Tijuana and had carried out studies in the past with monetary remuneration. In contrast, Mueller et al.<sup>40</sup> report much slower recruitment of IDUs in Las Cruces, NM using RDS, despite similar methodology and the same eligibility criteria.

The sample seroprevalence of HIV was relatively low in both cities. HIV-1 seroprevalence in IDUs in Tijuana recruited through RDS was similar to that found in IDUs studied by Güereña-Burgueño et al.<sup>41</sup> in the early 1990s and in a study by Magis-Rodriguez et al.<sup>42</sup> in 2003 that used time-location sampling methods, although the absolute number of HIV-positive cases was too low to perform reliable RDS corrections. In contrast to HIV, syphilis prevalence was extremely high, especially in women and in Tijuana.

Unlike many adaptive sampling schemes<sup>43</sup> in which the sampling process is controlled by the investigator, RDS enables the study subjects to control the sampling process. While this facilitates the recruitment process, it makes statistical inference more difficult. We found that estimates of syphilis seroprevalence were extremely sensitive to modeling assumptions. First, as recruitments between low-

TABLE 4. RDS-corrected estimates of syphilis seroprevalence, using the raw transition data, and adjusted degrees

City	Tijuana				Ciudad Juárez			
	Male		Female		Male		Female	
Syphilis antibody	Syphilis RPR−	Syphilis RPR+	Syphilis RPR−	Syphilis RPR+	Syphilis RPR−	Syphilis RPR+	Syphilis RPR−	Syphilis RPR+
n	174	20	5	8	158	3	8	2
Sample (%)	84.16	9.66	2.42	3.86	92.4	1.75	4.68	1.17
Equilibrium (%)	84.27	9.8	2.11	3.82	92.3	1.75	4.46	1.46
Recruitment weight	1	1.01	0.87	0.99	1	1	0.95	1.25
Degree	54.96	64.8	30	59.12	47.54	19.33	45.11	15
Adjusted degree	14.63	8.45	18.29	4.89	14.67	14.4	18.62	13.33
Adjusted degree weight	0.86	1.67	0.37	2.75	0.98	1	0.71	3.16
Adjusted sampling weights	0.86	1.69	0.32	2.72	0.98	1	0.67	3.95
Adjusted population (%)	72.37	16.35	0.77	10.5	90.47	1.75	3.16	4.62

TABLE 5. Sensitivity of the estimated population prevalence of syphilis antibody among men and women to model assumptions

Model assumptions	Population prevalence							
	Tijuana				Ciudad Juárez			
	Male		Female		Male		Female	
Transition matrix	Syphilis RPR−	Syphilis RPR+	Syphilis RPR−	Syphilis RPR+	Syphilis RPR−	Syphilis RPR+	Syphilis RPR−	Syphilis RPR+
Unadjusted	84.79	9.34	2.08	3.8	80.61	3.76	3.76	11.86
Unadjusted	72.37	16.35	0.77	10.5	90.47	1.75	3.16	4.62
Smoothed	84.81	8.98	2.75	3.45	87.94	3.97	4.69	3.41
Smoothed	73.35	15.94	1.04	9.67	93.28	1.74	3.72	1.25

Smoothed transition matrices are obtained by using the fitted values from the best fitting model in Table 3. Adjusted network sizes are obtained as described in Salganik and Heckathorn.<sup>33</sup>

frequency groups are relatively rare, estimates of the recruitment rates may be biased. Smoothing these estimates using a statistical model can lead to different estimates. Although simulations and analytical results show that RDS-based estimates are unbiased in large populations, errors in RDS based estimates may be so high for small populations and/or low frequencies of groups as to render the use of RDS impractical.<sup>33</sup> Secondly, the estimated prevalence of syphilis was sensitive to the assumption of how inclusion probability depends on reported network size. Estimates of network size may well have been different had we asked “How many people do you *currently* know by name or street name that inject drugs?” Estimating group-level network sizes is compromised by high variances, the small size of some of the subpopulations, and the poor ability of individuals to estimate the size of their personal networks.<sup>44</sup> Although RDS controls for differences in network sizes, a sampling bias long known to be inherent in chain-referral samples, it is important that this information is as accurate as possible. It might be argued that prior to the advent of RDS, there was little incentive to accurately measure relative network sizes in epidemiological studies; given that this information plays a crucial role in the post-stratification process of RDS, we encourage further research to determine how best to collect this information accurately.

RDS also has some inherent limitations in terms of inferences that can be drawn from the data<sup>45</sup>: it does not generate estimates of the absolute size of the population, only proportions, and it exploits social ties between individuals, limiting what one can conclude about sexual or drug-injecting networks from RDS data. Furthermore, without comparison of RDS to other types of sampling, we cannot conclude that obtaining a sample through RDS gives us a more representative sample than other methods. Nevertheless, RDS, or a modified version thereof, has the potential to efficiently recruit hidden populations such as IDUs, and creates avenues through which interventions can reach members of these populations. In the context of this study, prevention and treatment of syphilis is clearly an important public health concern.

## ACKNOWLEDGEMENTS

The authors gratefully acknowledge support from the National Institute on Drug Abuse through grants DA09225-S11 and DA019829, donor support for the Harold Simon Chair in International Health and Cross-Cultural Medicine, the UCSD Center for AIDS Research (P30 AI36214), and training grants supporting Dr. Brouwer (T32 AI07384 and K01 DA20364). In addition, Dr. Frost is supported by the National Institute of Allergy and Infectious Disease (grants U01 43638, R01 AI47745, and R01 AI57167), and by the National Institute on Drug Abuse, as part of the Sexual Acquisition and Transmission of HIV Cooperative Agreement Program (grant DA17934). We are extremely thankful to the staff of Proyecto El Cuete, CIRAD, COMUSIDA and Programa Compañeros, and to Dr. Peter Hartsock, Dr. Doug Heckathorn and Mr. Cyprian Wejnert.

## REFERENCES

1. Brouwer KC, Case P, Ramos R, et al. Trends in production, trafficking and consumption of methamphetamine and cocaine in México. *Subst Use Misuse* 2006;41:707–727.
2. Beyrer C, Razak MH, Lisam K, Chen J, Lui W, Yu XF. Overland heroin trafficking routes and HIV-1 spread in south and south-east Asia. *AIDS*. 2000;14:75–83.

3. Bucardo J, Brouwer KC, Magis-Rodriguez C, et al. Historical trends in the production and consumption of illicit drugs in Mexico: implications for the prevention of blood borne infections. *Drug Alcohol Depend.* 2005;79:281–293.
4. Freeman RC, Rodriguez GM, French JF. A comparison of male and female intravenous drug users' risk behaviors for HIV infection. *Am J Drug Alcohol Abuse.* 1994;20: 129–157.
5. Kral A, Bluthenthal RN, Lorvick J, Gee L, Bacchetti P, Edlin BR. Sexual transmission of HIV-1 among injection drug users in San Francisco, USA: risk-factor analysis. *Lancet.* 2001;357:1397–1401.
6. Strathdee SA. Sexual HIV transmission in the context of injection drug use: implications for interventions. *Int J Drug Policy.* 2003;14:79–81.
7. Strathdee SA, Sherman SG. The role of sexual transmission of HIV infection among injection and non-injection drug users. *J Urban Health.* 2003;80:iii7–iii14.
8. Friedman SR, Ompad DC, Maslow C, et al. HIV prevalence, risk behaviors, and high-risk sexual and injection networks among young women injectors who have sex with women. *Am J Public Health.* 2003;93:902–906.
9. Benotsch E, Somlai AM, Pinkerton SD, et al. Drug use and sexual risk behaviours among female Russian IDUs who exchange sex for money or drugs. *Int J STD AIDS.* 2004;15:343–347.
10. Perngman P, Celentano DD, Kawichai S. Sexual risks among southern Thai drug injectors. *AIDS Behav.* 2004;8:63–72.
11. UNAIDS. *Report on the Global HIV/AIDS Epidemic.* Geneva, Switzerland: UNAIDS; 2004.
12. CENSIDA. Panorama Epidemiológico del VIH/SIDA e ITS en México. México City: Secretaria de Salud; 2004. Available at <http://www.salud.gob.mx/conasida>. Accessed on January 24, 2006
13. del Rio CC, Sepulveda J. AIDS in México: lessons learned and implications for developing countries. *AIDS.* 2002;16:1445–1457.
14. Magis-Rodríguez C, Marques LF, Touze G. HIV and injection drug use in Latin America. *AIDS.* 2002;16(Suppl 3):34–41.
15. Viani RM, Calderon JR, Van Pratt C, et al. HIV prevalence during pregnancy in Tijuana, Baja California, Mexico. *AIDS.* 2003;17:1113–1114.
16. Viani RM, Araneta MRG, Ruiz-Calderon J, et al. Perinatal HIV Counseling and Rapid Testing in Tijuana, Baja California, Mexico: seroprevalence and correlates of HIV infection. *J AIDS.* 2006;41:87–92.
17. Valdez A, Cepeda A, Kaplan CD, Codina E. Sex work, high-risk sexual behavior and injecting drug use on the U.S.–México border: Ciudad Juárez, Chihuahua. Houston, TX: *Office for Drug and Social Policy Research*, Graduate School of Social Work, University of Houston, Texas, USA; 2002.
18. Patterson TL, Semple SJ, Fraga M, et al. Comparison of sexual and drug use behaviors between female sex workers in Tijuana and Ciudad Juárez, Mexico. *Subst Use Misuse.* 2006 (in press).
19. Wasserheit JN. Epidemiological synergy: interrelationships between human immunodeficiency virus infection and other sexually transmitted diseases. *Sex Transm Dis.* 1992;19:61–77.
20. Padian NS, Shiboski SC, Glass SO, Vittinghoff E. Heterosexual transmission of human immunodeficiency virus (HIV) in northern California: results from a ten-year study. *Am J Epidemiol.* 1997;146:350–357.
21. Centers for Disease Control and Prevention. HIV prevention through early detection and treatment of other sexually transmitted diseases—United States. *Morb Mort Wkly Rep.* 1998;47:1–24.
22. Blocker ME, Levine WC, St Louis ME. HIV prevalence in patients with syphilis, United States. *Sex Transm Dis.* 2000;27:53–59.
23. Conde-Glez CJ, Calderon E, Juarez-Figueroa L, Hernandez-Avila M. Historical account of venereal diseases in Mexico. *Genitourin Med.* 1993;69:462–466.

24. Valdespino-Gomez JL, Garcia-Garcia M, del Rio-Zolezzi A, Loo-Mendez E, Magis-Rodriguez C, Salcedo-Alvarez RA. Epidemiologia del SIDA/VIH en Mexico. *Salud Publica Mex.* 1995;37:520–524.
25. Uribe-Sals F, del Río C, Conde-Glez CJ, et al. Prevalence, incidence and determinants of syphilis in female commercial sex workers in México City. *Sex Transm Dis.* 1996; 23:120–126.
26. Uribe-Salas F, Hernández-Avila M, Conde-Glez CJ, et al. Low prevalence of HIV infection and sexually transmitted diseases among female commercial sex workers in México City. *Am J Public Health* 1997;87:1012–1015.
27. Zamilpa-Mejía LG, Uribe-Salas F, Juárez-Figueroa L, Calderón-Jaimes E, Conde-González CJ. Prevalence and factors associated with syphilis and genital herpes in two female population groups. *Salud Publica Mex.* 2003;45(Suppl 5):S617–S623.
28. Heckathorn D. Respondent driven sampling: a new approach to the study of hidden populations. *Soc Probl.* 1997;44:174–199.
29. Heckathorn D. Respondent-driven sampling II: deriving valid population estimates from chain-referral samples of hidden populations. *Soc Probl.* 2002;49:11–34.
30. Anderson TW, Goodman LA. Statistical inference about Markov chains. *Ann Math Stat.* 1957;28:89–110.
31. Agresti A. *Categorical Data Analysis*. John Wiley & Sons, Inc., Hoboken, USA; 2002.
32. Akaike H. A new look at the statistical model identification. *IEEE Trans Automat Contr.* 1974;AC-19:716–723.
33. Salganik MJ, Heckathorn DD. Sampling and estimation in hidden populations using respondent-driven sampling. *Sociol Method.* 2004;34:193–240.
34. Ihaka R, Gentleman R. R: a language for data analysis and graphics. *J Comp Graph Stat.* 1996;5:299–314.
35. Broadhead RS, Heckathorn DD. AIDS prevention outreach among injection drug users: agency problems and new approaches. *Soc Probl.* 1994;41:473–495.
36. Broadhead R, Heckathorn D, Weakliem D, et al. Harnessing peer networks as an instrument for AIDS prevention: results from a peer-driven intervention. *Public Health Rep.* 1998;113(Suppl 1):42–57.
37. Heckathorn D, Sergeyev B, Oparina T, et al. HIV prevention in Yaroslavl, Russia: a peer-driven intervention and needle exchange. *J Drug Issues* 1999;29:777–804.
38. Heckathorn D, Broadhead RS, Anthony DL, Weakliem DL. AIDS and social networks: prevention through network mobilization. *Sociol Focus.* 1999;32:159–179.
39. Heckathorn DD, Semaan S, Broadhead RS, Hughes JJ. Extensions of respondent-driven sampling: a new approach to the study of injection drug users aged 18–25. *AIDS Behav.* 2002;6:55–67.
40. Mueller MR, Robins AE, Foster LN, et al. Respondent-driven sampling to recruit active injection drug users in a nonurban setting: challenges and lessons learned—New Mexico, 2005. *J Urban Health* (submitted).
41. Güereña-Burgueño F, Benenson AS, Sepulveda-Amor J. HIV-1 prevalence in selected Tijuana sub-populations. *Am J Public Health.* 1991;81:623–625.
42. Magis-Rodriguez C, Brouwer KC, Morales S, et al. HIV prevalence and correlates of receptive needle sharing among injection drug users in the Mexican–U.S. border city of Tijuana. *J Psychoact Drugs.* 2005;37:333–339.
43. Thompson SK, Frank O. Model-based estimation with link-tracing sampling designs. *Surv Methodol.* 2000;26:87–98.
44. Kochen M. *The Small World*. Norwood, New Jersey: Ablex; 1989:176–210.
45. Heimer R. Critical issues and further questions about respondent-driven sampling: comment on Ramirez-Valles, et al. (2005). *AIDS Behav.* 2005;9:403–408.
46. Heckathorn D. Extensions of respondent-driven sampling: dual component sampling weights. Paper presented at the RAND Statistics Seminar Series, Santa Monica, CA, December 8, 2005.